Deconstructing Explanation by Mechanism

by Hannu Ruonavaara
University of Turku

Sociological Research Online, 17 (2) 7
<<http://www.socresonline.org.uk/17/2/7.html>>
10.5153/sro.2504

Received: 17 Oct 2011 Accepted: 18 Jan 2012 Published: 31 May 2012

Abstract

Critical realism and analytic sociology support what can be called explanation by mechanism (EbM). EbM consists of three distinct theses. The elaboration thesis says that mere observation of interdependencies cannot be considered a full explanation because it does not give a description of how the proposed causal connection comes about. The mechanism thesis claims that elaboration must contain an account of the generative/causal mechanism at work. The actionalism thesis cannot be placed within all versions of EbM because it states that an account of a mechanism should describe (1) the situation of the agents and (2) the actions the agents undertake on the basis of their situation. The paper argues that both the elaboration thesis and the actionalism thesis are of crucial importance. However, elaboration should not be equated only with an account of mechanism(s) because a 'theory' of the agent (agent-image) and an account of the relevant context(s) are needed.

Keywords: *Agent-Image, Analytical Sociology, Causality, Critical Realism, Explanation, Intention, Social Mechanism*

Introduction

1.1 'Mechanism' is one of the metaphors social scientists use when trying to comprehend the nature of the social world, and according to Hernes (1998, 74), it is a very powerful one. In the social sciences it is quite usual to speak of market mechanisms, mechanisms of exclusion, mechanisms of power, etc. The metaphor of mechanism is considered to be one of the root metaphors of sociology, which are sets of assumptions about what kinds of things there are in the social world, how they work, how they connect and how we can know about them (Brown 1977, 125). In other words, root metaphors are taken-for-granted conceptual devices that sociologists often use without much reflection. This is also the case with the metaphor of mechanism; it has been routinely used with reference to the workings of some social system. In recent times there has been much theoretical interest in the concept of mechanism. "Mechanisms are 'in the air' these days", writes Philip Gorski (2009, 151; see also Hedström & Ylikoski 2010, 50). But what are we talking about when we use the word 'mechanism' in sociology?

1.2 'Mechanism' can be a descriptive term, like when speaking 'mechanism of government', but it can also have explanatory content. For example, certain outcomes are there to be observed because there is a market mechanism at work. Thus, the 'mechanism' works as expected and produces the observed outcomes. However, the term becomes more interesting in sociology when things *do not* function as expected. In some situations there can be an *intervening* mechanism that alters the expected outcome. In recent times the notion of mechanism has become important in discussions concerning sociological explanation.

1.3 The standard conception of causal explanation in sociology is still the *covering law* theory of explanation, famously explicated (though not invented) by Carl Hempel in the 1940s. According to this view, explanation is the deduction of a proposition stating what is to be explained (explanandum) from a set of propositions (explanans) stating a universal or statistical law and initial conditions. This *deductivist* view of explanation has long been subject to criticisms which have led to various revisions of the original model. According to an alternative view, explanation is rather a process or an account by which something (explanans) is made intelligible and it does not have any standard logical form (see Bhaskar 1986, 56; see also Bhaskar 1979 and Harré 1988). Such a discourse might be formulated in the covering

law format but this format is neither necessary nor sufficient for it to be an explanation. One alternative to the covering law model is what Elster has called *explanation by mechanism* (Elster 1989b, 3; also Elster 1998, 47). Below I shall refer to this conception with the acronym EbM.

1.4 Most forcefully EbM has been propagated by sociologists associated with critical realism (see e.g. Keat & Urry 1982; Sayer 1984; Pawson 1989; Manicas 2006)^[1]. However, critical realists are not the only social scientists supporting EbM. There is a loose school of social researchers, more or less inspired by rational action theory and Robert Merton's version of functionalism (see Merton 1968), which considers accounting for the social mechanisms at work in social reality a central task in the business of social explanation. The early manifesto of this school is the anthology edited by Hedström and Swedberg in 1998, *Social Mechanisms. An analytical approach to social theory*. More recently Hedström started to call this strand of social scientific thinking *analytical sociology* (Hedström 2005; see also Manzo 2000). I will use this name to refer to this group of social scientists and their version of EbM – though the views of the scholars that wrote in the 1998 manifesto do not coincide in all aspects.

1.5 Nor do the views of critical realists and analytical sociologists coincide, even in some of the most central postulates of the two approaches. For example the central thesis of critical realists – that social reality is stratified into ontologically distinct levels that cannot be reduced to the other levels – is severely criticised by Hedström (2005, 70-74). Moreover, his, and other analytical sociologists,' methodology is individualist, whereas critical realists tend to dismiss methodological individualism as ignoring the 'fact that individuals and institutions operate in many different structures' (Sayer 1984, 120). However, these two strands of social research seem to share a similar view of what explanation is all about. Moreover, there are other theoretical orientations supporting EbM than those associated with either critical realism or analytical sociology. For example, Gross has provided a pragmatist interpretation of EbM (Gross 2009). It seems that *EbM is a conception of explanation that can be accepted while holding rather different methodological and meta-theoretical convictions* (see also van Parijs 1981, 14). It is a general view of explanation that does not carry a lot of theoretical or methodological baggage.

What exactly is EbM?

2.1 The basic idea of EbM is that to *really* explain something is to describe the mechanisms that mediate causes and effects and to show that this explanation actually holds true. However, a problem concerning the use of the term 'mechanism' is that there seems to be only a 'family resemblance' between all the different uses of the term. In his admittedly limited review on the literature, Mahoney found as many as 24 different definitions of mechanism (Mahoney 2001, table 1). Falleti and Lynch list 14 different definitions (Falleti and Lynch 2009, table 1) whereas Hedström and Ylikoski list 9 definitions (Hedström and Ylikoski 2010, table 1).

2.2 Nevertheless, although there are many definitions, absolute confusion about what 'mechanism' means does not reign. The different uses of the term share certain assumptions. First, mechanisms are regular *patterns* of specific kinds of events and states of affairs. In the case of social mechanisms (which are the primary interest in this paper) these are actions and interactions, possibly also social relations. Secondly, mechanisms are *causally productive*: they bring about the outcomes that we are interested in. Therefore, demonstrating the mechanism at work in the state of affairs studied will also provide an explanation of how that state of affairs came to be as it is. Thirdly, the idea of mechanism *implies* the possibility that the pattern discovered in one context can be discovered in others. Mechanisms are *portable*, as Falleti and Lynch put it (2009, 19). There is no point in talking about mechanisms if what is discovered is a singular trajectory of events producing other events without any generalisable element. However, accounts of social mechanisms are not general theories for the reason that while they are somewhat general, they are not universally general in the way general theories are supposed to be.

2.3 Assuming that there is some agreement on what 'mechanism' means, there is a problem concerning what kinds of entities mechanisms are considered to be. There is an important difference between critical realists and *some* analytical sociologists concerning the ontological status of explanatory mechanisms. Critical realists very clearly assume that mechanisms are entities that exist, are part of reality. This is also Hedström's view (see 2005, 38 and Hedström & Ylikoski 2010, 52). However, not all writers associated with analytical sociology necessarily think so. In *Social Mechanisms* Gudmund Hernes writes about the two worlds of social scientists, the real world and the conceptual and idealized virtual world that social scientists create to make sense of the real world. For Hernes, causal mechanisms are a part of that simulated world, not the real one (Hernes 1998). This does not suggest a realist theory of social science but an instrumentalist one, one where theories and models are arbitrary constructs that facilitate the discussion of empirical observations^[2] (see e.g. Sayer 1984, 70-71).

2.4 An instrumentalist approach to EbM leads easily to an emphasis on models rather than mechanisms. When Hernes discusses mechanisms more concretely, he completely forgets the word 'mechanism' and talks only about theoretical models. It seems that 'mechanism' is not needed. A possibility is that 'mechanism' can be used as a name for a particular type of model. However, Hernes does not indicate that he would use the term in this sense. Thus, in Hernes' usage the concept of mechanism actually becomes redundant. The alternative view is that the productive constellations of actions and interactions that the mechanism concept is used to refer to actually exists. There *are* mechanisms in reality and these can be approached by theoretical models. This does not mean that the accounts of mechanisms are one to one descriptions of reality. Rather they are helpful analytical models or ideal types (Hedström & Swedberg 1998, 13) that represent the social world in a radically simplified manner, focusing on those aspects of the situation that are considered most essential (Hedström & Ylikoski 2010). In Hedström's terms they are not descriptively complete accounts of reality, but are based on assumptions and statements that are not arbitrary and reflect what we believe to be true in the world.

2.5 Whatever the differences there are between the critical realist's and analytical sociologist's view of mechanisms, there seems to be the broadly similar logic of EbM in both approaches (compare Outhwaite 1987, 33 and Hedström & Ylikoski 2010, 63). Hence, if we assume that a social state of affairs is a term for all those things we would wish to explain in sociology, we can say that in both approaches the recipe for explanation involves the following:

1. Identify a social state of affairs needing explanation.
2. Imagine a mechanism or several alternative mechanisms that would produce the state of affairs if it or they were in operation in the situation.
3. Derive testable propositions for the operation of each of the imagined mechanisms.
4. Establish evidence that could be used to evaluate whether the propositions are correct or not.
5. Eliminate those accounts of mechanism that are not supported by evidence and accept the account that is best supported by evidence.

2.6 As can be seen, this is an account that does not radically depart from a basically Popperian view of social and other sciences. There is an element of hypothesis-testing in it. The ideas about mechanisms must be corroborated in some way. For critical realists this may be through 'experimental activity' (Outhwaite 1987, 33), statistical analysis (Pawson 1989), 'intensive' (=qualitative) research (Sayer 1984) or historical analysis (Lloyd 1986). The analytical sociologists add agent-based simulation to this list of methods in social science (Hedström & Ylikoski 2010, 62-64). Whatever the chosen method, the logic remains broadly the same and is encapsulated by Hedström and Ylikoski (2010, 53): 'The empirical evidence turns a possible mechanism into a plausible mechanism and may eventually lead to the identification of the actual mechanism'. The quotation points first to the double meaning that 'mechanism' has: it refers both to the researcher's hypothesis of the constellation of forces that might explain the outcome that is being examined (the possible mechanism) and to the real constellation of forces that actually produce the outcome (the actual mechanism). There is no direct route to actual mechanisms, so their existence is an inference based on mechanism hypotheses that are supported by empirical evidence; making the possible mechanism into a plausible one.

2.7 The paper so far has been rather abstract. The reader may ask: what exactly counts as a mechanism? Critical realists are generally not very good at giving examples of what they mean by mechanism. For instance, Marx's 'tendency of rate of profit to fall', which Marx himself considered a social scientific *law*, is cited by Sayer as an example (Sayer 1984, 106). Rom Harré's favourite example comes from natural science; though the mechanism in question, natural selection, can arguably be applied to a variety of sociological problems (see e.g. Runciman 2000). According to Harré, making more observations and collecting more specimens could not answer the puzzle about the change of species that Darwin had discovered. Therefore, he had to imagine a mechanism that would 'endow it with existential plausibility' (Harré 1988, 140; see also Harré, Clarke & De Carlo 1985, 4344). Analytical sociologists tend to search for examples from social psychology, such as cognitive dissonance. However, other good examples can be found from Robert Merton's reservoir of sociological insights, for example the self-fulfilling prophecy, while Thomas Schelling's models on micromotives and macrobehavior provide more recent examples. My chosen example comes from housing studies, although, and similar to many other examples, the author of the papers I am citing did not consistently use the term 'mechanism' to refer to the constellation of actions and interactions that he discusses. With this example, I also want to highlight a broader explanatory use of mechanism than the textbook version outlined above.

2.8 The Swedish housing researcher Bo Bengtsson has investigated the well-known fact that residents are difficult to organise into a joint action group. His interpretation of why there is so little cooperative activity between tenants in residential areas is a variant of the free-rider problem (Bengtsson 1998). Typically this type of cooperation aims to make the environment better and to enhance the enjoyment of the residential area. The results of this activity benefit all residents, not just those who have participated in the activities. Therefore a self-serving resident should abstain from cooperative activity and let others do the work. However, when many residents think like that, there is no cooperative activity and the residents cannot receive any benefit. This Bengtsson calls the *tenants' dilemma*. The dilemma is that everyone would benefit from the cooperative activities but participating is against everyone's immediate advantage. The tenants' dilemma can be thought of as a certain kind of social mechanism. It explains the known fact that it is fairly difficult to activate the residents to participate in cooperative activities in their environment, although the general attitude towards cooperative activities is positive.

2.9 The tenants' dilemma is a theoretical interpretation of a known empirical fact, but Bengtsson's team was not interested in verifying it empirically. Instead they wished to examine situations where the dilemma has been overcome. It so happens that in Sweden there are a number of case studies concerning tenant co-operation in residential areas. In the research 26 previously studied residential areas were investigated 'in order to trace the mechanisms behind the rise and possible fall of collective action' (Bengtsson 2001, 179). Then the successful cases where co-operation was sustained were put under closer analysis. Bengtsson emphasises that the research of the team is historical and qualitative and adds '[a]s always when we try to learn something from history, we tend to identify mechanisms rather than count cases' (Bengtsson 2001, 179).

2.10 In interpreting the cases Bengtsson draws on Jon Elster's ideal type of three norms of co-operation: everyday Kantianism, the norm of reciprocity and the norm of (local) utilitarianism. These imply different prescriptions for co-operation:

The norm of everyday Kantianism says: "If I don't co-operate, why should anyone else?" The norm of reciprocity says: "If others co-operate, why shouldn't I?" Finally, the prescription from the norm of local utilitarianism reads: "I take part if it is needed and if I can contribute to the collective good in my estate." (Bengtsson 2001, 181.)

2.11 The central finding of the study is that local utilitarianism "has been the central mechanism behind the institutionalization of collective action in our estates" (Bengtsson 2001, 182). Note Bengtsson's shift from norm to mechanism. Later it is specified that the *development* of the norm of local utilitarianism is actually the mechanism (Bengtsson 2001, 185) rather than the norm itself. For the norm to emerge, certain conditions have to be present: the individual must have time and resources to participate, her/his contribution is needed and the collective benefit for co-operation must exceed the collective cost (Bengtsson 2001, 182). Meeting these conditions creates the mechanism that produces in people the obligation to participate and consequently sustains co-operation.

2.12 Bengtsson's account actually singles out four different explanatory social mechanisms, although only three are investigated empirically. The starting point in this research, the tenants' dilemma, is wholly a *theoretical* mechanism in the sense that it is not put to an empirical test. It rather poses the question to be investigated: *why is there*, rather than *why is there not*, co-operation? This theoretical mechanism can be seen as derived from a well-established theory that has been tested elsewhere (with varying results). The other mechanisms are tested according to the procedure laid out in my explication of the logic of EbM. However, instead of experiments or simulations, multi-method case studies were conducted. The way that the idea of mechanism is used in this study points to a more flexible use than the standard version of EbM would suggest.

2.13 The example also highlights the theory-laden nature of mechanism *hypotheses*. All of Bengtsson's hypothetical mechanisms were more or less constructed in terms of a meta-theory of action that conforms to a sociologised, contextual rational action theory (cf. Somerville & Bengtsson 2002). Stinchcombe has characterised mechanisms as theories within theories. Kiser and Hechter have even argued – in the context of a debate on the methodology of comparative-historical sociology – that mechanisms can *only* come out of general theories, and not from inferences from empirical observations (see Kiser & Hechter 1991). However, as has been pointed out as a critique of this position, there is no reason why mechanisms could not be constructed on theoretical ideas arising from empirical work, rather than axiomatic theorising that makes theory general (see e.g. Quadagno & Knapp 1992).

Deconstructing EbM

3.1 What does it mean that an explanation is or should be an account of the generative or causal mechanism? To think of the above question I will borrow some terminological tools from Philippe Van Parijs. According to him, causal explanation consists of two parts. First one has discovered a *causal link* between two entities, e.g. a regular correlation between, say, gender and a positive attitude towards foreigners. Unfortunately, this is not enough. One has to reach *causal understanding* which 'consists in imagining a plausible mechanism through which the empirical fact to be explained is brought about, produced, caused' (Van Parijs 1981, 14). Thus, in the example, this would require an account of how gender might be associated with a favourable attitude towards foreigners. Van Parijs' thesis is that a causal explanation is not adequate without causal understanding. He points out that one can arrive at this view from a different epistemology than the (external) realism of the social realists. For those who hold a conventionalist view of science (like Kuhn, Hanson or Habermas) explaining something means providing an account of it that will be accepted as intelligible by the scientific community. Intelligibility is not guaranteed by merely stating the causal link but by providing an understandable account of the underlying process, that is, the generative mechanism. I will call such accounting that provides a deeper understanding of the causal link, an *elaboration* of the explanation (the term is from Cohen 1980). Elaboration is an account that opens the 'black box' between cause and effect and thus provides causal understanding. Now the idea of explanation by mechanism can be broken down into two theses.

The elaboration thesis. The mere observation of interdependencies between factors cannot be considered a full explanation of the phenomenon to be explained. What is needed is an elaboration of the connection that involves telling a causal story of *how* that connection comes about.

3.2 In the first thesis there is an observed interdependency between certain phenomena and an intuitive understanding that there is a causal link between them. To be precise, to speak of a causal link is not possible without some kind of 'theory' about it but here the theory is not worked out into a full explication of how the link comes about. The thesis actually says that a 'complete' explanation should provide a description giving a causal understanding of the link in question.

The mechanism thesis: Elaboration amounts to a detailed description of the generative/causal mechanism at work.

3.3 The mechanism thesis says that the account of a state of affairs, which provides a causal understanding of that state, should be constructed in terms of social and other mechanisms. In other words, 'mechanism' is the metaphor, concept and analogue that provides adequate causal understanding. This means that mechanism is something more than just 'a systematic set of statements that provide a plausible account of how *I* and *O* are linked to one another' (Hedström & Swedberg 1998, 7). The set of statements in the quotation define what has been termed elaboration in this paper. In principle, an elaboration can be something other than a description of mechanism – assuming that the word has a more specific meaning. If any elaboration is a mechanism, there is no specific content to the term mechanism. This is a point I shall return to later.

3.4 Many conceptions of EbM also seem to entail a third thesis that I shall call the *actionalism thesis*. By actionalism van Parijs refers to the belief that all adequate explanations must be reduced to the level of agents and their situations (van Parijs 1981, 18-23; see also Popper 1979, 178-180).

The actionalism thesis: the elaborations and mechanisms should account for (1) the situation of the agents involved and (2) the actions the agents undertake on the basis of their situation

3.5 In the work of analytical sociologists actionalism is the antidote against 'variables sociology' making causal claims entirely on the basis of observations concerning correlations (e.g. Hedström & Swedberg 1998, 9-11; Räsänen 2003, 74-75). Correlations need accounting for if they are to be made into causal claims.

3.6 Elster and other analytical sociologists tend to think that accounting for the mechanism(s) at work reduces the explanation to the level of actors. What makes the mechanism intelligible is that it accounts for how the actors produce the effect to be explained. There is an obvious individualist emphasis in analytical sociologists' writings (see Abbott 2007, 3-6), even though analytical sociologists would like to qualify their position as *structural individualism*, which acknowledges the 'explanatory importance of relations and relational structures' (Hedström & Ylikoski 2010, 60). However, actionalism is not a necessary element of EbM. Accounts of mechanisms can, in principle, be constructed without reducing the explanation at the level of individual or collective actors and their choices in situations. I shall return to the point of whether this is sensible or not in the concluding section of this paper.

Assessing the theses

4.1 I believe the elaboration thesis to be a very important one, and one that provides the working social researcher with clear methodological advice: when thinking that there is a causal link between some factors, always try to figure out how that link is brought out. When making claims about causal links, we should be able to provide elaborations. For example, the often observed link between higher middle class social background and entry into higher education is usually summarised by saying that social class explains educational outcomes. However, it is not social background that provides the explanation but a bundle of factors, like economic resources, behavioural dispositions, the ambitions typical of the different classes, being able to conceive of opportunities, self-confidence, etc. Those factors are distributed in different ways among people of different social backgrounds and they amount to a consistent link between social background and educational career. When we read claims about causal relations we should always ask: how does this come about?

4.2 Why then is it the case that unelaborated statistical statements as well as other statements of causal links are often presented to us as adequate causal explanations? When reading such statements we often have intuitive conceptions of the factors and the connections that provide the actual understanding of the link. Therefore unelaborated causal claims can sometimes be accepted as full explanations. For example, there is such a link between gender and attitudes towards foreigners that shows that women tend to have more favourable attitudes than men. We can imagine a number of reasons why this should be so. However, we should not take such an explanation as self-evident but require that the link is elaborated. Without such an elaboration we cannot adequately assess the validity of the explanation. It is also problematic that we can perhaps provide many intuitively satisfactory, even contradictory reasons for women to have more favourable attitudes towards foreigners than men. Only by assessing the plausibility of the reasons and applying some kind of empirical test to them – where we compare different kinds of elaborations – can we approach real causal understanding.

4.3 Accepting the elaboration thesis does not necessarily entail accepting the mechanism thesis. It is not absolutely necessary to understand elaborations as mechanisms. In an early paper, one of the social mechanism sociologists, John Goldthorpe wrote about *action storylines* that make links between macro-sociological variables understandable at the level of action (cited in Wallis & Bruce 1986, 34–35). Based on Goldthorpe, Wallis and Bruce state that explanation resides in the action storylines that make the connection between the variables understandable, but not in the connection itself. This is a statement that is wholly in accordance with the elaboration thesis, but does not use the terminology of mechanisms at all. Without having access to the original text it is hard to say whether the idea of trying to grasp 'frequently observed ways in which things happen' (Elster 1989a, viii), which is central to the mechanism thinking, is built into the idea of action story-lines or not. Sociology can be thought of as being exclusively interested in recurrent storylines rather than singular ones.

4.4 Is 'mechanism' the right kind of metaphor and analogue to account for what elaborations are all about? After all, it is a word that refers to objects rather than social life, and all of its connotations may not be what we want to accept about social life. The word mechanism refers to machinery, the parts of a larger machine. According to Brown, 'mechanism' was conceived as a metaphor in connection with social life in the beginning of the modern period and it was intimately connected with the then emerging scientific worldview (Brown 1977, 139). Is it perhaps a metaphor that carries the baggage of a perspective that has become outdated since its conception? In the time of digital technology should we create new kinds of metaphors to account for the social world? Does the metaphor lead us to think of society as clockwork? What I am hinting at is the suspicion that the metaphor of mechanism makes our view of society too mechanistic at a time when our theories emphasise the increasing non-mechanistic unpredictability of society. ^[3]

4.5 The obvious counterargument is that in social sciences 'mechanism' is a technical term that has nothing to do with the idea that society works as clockwork. But is this really so? In social research words we use to describe ideas often carry baggage that is undesirable. This is even more pronounced when they are a part of the language that the people we study use. To avoid the problem, the researcher may want to develop neologisms that are not part of the everyday language. However, that causes problems of its own. For example, W.G. Runciman wants to use the word 'syntact' to refer to certain kinds of

structurally defined groups of actors (e.g. Runciman 2000) – but the term is much more difficult to understand than the familiar terms "class" and "status group" that Runciman also uses.

The problem with intentional action

5.1 The idea of mechanism is somewhat problematically associated with the idea of intentional action. When we talk about social mechanisms we often think of sets of factors that work so that the outcomes are something other than those the individual actors intended them to be. Social mechanisms work 'behind the backs' of actors to produce outcomes that are not especially intended by them and sometimes work against their intentions. For example, one of the analytical sociologists' favourite examples, Thomas Schelling's model of social segregation (see Schelling 1971), is just of this kind. Schelling argues that the residential choices of actors that do not have especially segregationist attitudes may produce a strongly segregated residential structure. Only a very weak preference for living with people that are similar to oneself is enough to produce segregation if sufficient numbers of people have such a preference. What Hedström and Bearman (2009, 12) find illuminating in this example is that the social mechanism postulated produces aggregate outcomes that are unintended by the actors, leading to cases where 'small and seemingly trivial differences at the micro level can make huge differences at the macro level'. It seems that, for analytical sociologists, interesting social mechanisms are those that produce non-intentional outcomes, mechanisms that work behind the backs of actors.

5.2 Let us take an example of the tension between ideas of social mechanism and intentional action. In his semi-classic study of the American temperance movement, *The Symbolic Crusade* (1986/1963), Joseph Gusfield interpreted the mass support for the temperance movement in late 19th and early 20th century as a result of a middle class status panic. Identification with the ideals of temperance was for middle class supporters a political way to defend and promote the way of life and the prestige of the middle class. That was their main concern, not the high ideals of the movement. Roy Wallis (1979, 97-98) criticised this interpretation as poorly supported by empirical evidence. There was no evidence that the supporters would have felt their status position in the society threatened or that this feeling would have motivated their actions. Instead of the status politics interpretation, Wallis suggests an interpretation stressing the cultural difference between the American Protestant attitudes towards drinking and those of the new immigrant groups. People identifying with the American Protestant values felt their culture threatened irrespective of social position and were thus mobilised into temperance movement.

5.3 Gusfield's interpretation can easily be thought to be an explanation in terms of mechanisms. The explicit goals of the temperance movement are revealed to be a veil that hides not so very conscious covert goals. Hunt (1999) has written about an 'anxiety theory' that many social theories from Gusfield to Bourdieu use to account for different kinds of morally motivated action. That theory says, 'the occurrence and timing of some social phenomena is explained by reference to the presence of some elevated state of anxiety which elicits social or political responses by an identifiable group of social agents' (Hunt 1999, 509). With little paraphrasing, it can be said that Gusfield's explanation of the popularity of the temperance movement is one kind of *anxiety mechanism* – which is a social-psychological rather than a purely social mechanism. Another possibility is that there was a genuine social mechanism that unintentionally produced the outcome that the temperance movement promoted, i.e. the middle class status position, although its members were thinking they were acting on purely moral grounds. There might be elements of such an interpretation in Gusfield's book as he (contrary to what Wallis says) claims not to disregard the moral motivation behind the temperance movement.

5.4 If Gusfield's interpretation can be seen as an EbM, is that the case with Wallis' suggested alternative interpretation? If there is a mechanism, it is an utterly transparent one: people observed their values to be in danger and acted upon that observation. I think this kind of intentional, in a way rational, action is not what we think when we talk about mechanisms either social or otherwise. Nevertheless, such an account is explanatory and can provide a perfectly valid elaboration of a causal link. After all, rational actor theorists sometimes say that rational action is self-explanatory: if some action is observed to follow the canons of rational action, no other explanation is needed. To be precise, the theory postulates that people tend to act rationally, rather than emotionally, out of habit or to conform with social expectations. If there are other explanatory factors at work than social mechanisms, then the elaboration of a possible causal link is not equal to accounting for the social mechanism(s).

An alternative terminology

6.1 One way to discuss what elaboration is all about is to make a distinction between what I shall call *agent-images* and social and other mechanisms. There are many possible agent-images. The basic agent-image of all kinds of rational action theories (RATs), whether they are strict or loose, is that action is first and foremost explained by the rational pursuit of the agent's goals and objectives. While most other sociologies would just as well accept the view that human actors are predisposed to act rationally in the sense described by RAT, not all would be content with stating that this is the main defining character of human agency. We can contrast the RAT agent-image, for example, with a pragmatist one saying that human actors are first and foremost predisposed into acting out of habit, or with a functionalist agent-image that says that human actors tend to act in conformity with social norms. Furthermore, sociological approaches that do not take the intentional human actor as their starting point, like Michel Foucault's genealogy or Harrison White's network structuralism, have *implicit* agent-images for the simple, generally uncontested reason that human society happens only through human actions. Agent-images are theoretical constructs that cannot easily be proven right or wrong and RAT is a particularly good example of this. Practically *any* behaviour, however altruistic, can be explained as being targeted towards some hidden benefit (see e.g. Sen 2006). However, there are limits to how far one can go in postulating concealed benefits in whatever action.

6.2 Let's assume that RAT's agent-image is the one we want to work with. The rational actor can be seen as influenced by both sub- and supra-individual mechanisms. While the book edited by Hedström and Swedberg that laid down the foundations of analytic sociology's position on explanation was titled *Social Mechanisms*, some of its contributors like Elster and Gambetta (see Gambetta 1998) were actually more interested in the psychological and cognitive mechanisms affecting individuals' motivation: cognitive dissonance, adaptive preferences, wishful thinking, etc. These are mechanisms that make actors act in ways that are not rational in the sense specified in RAT. But they are not social mechanisms in the book's editors' definition of the term. In the introduction to their anthology Hedström and Swedberg write about *social mechanisms* concerning regularly observed, causally effective patterns of interaction between actors (Hedström & Swedberg 1998). That is what I suggest 'social mechanism' should be restricted to mean. Most of the examples of social mechanism, from self-fulfilling prophecy to tenants' dilemma and Schelling's segregation mechanism, are about the joint outcomes of individual actions and interactions, which are something other than those the actors individually intend. Secondly they can affect the parameters of the actors' choices: beliefs and preferences and modes of reasoning.

6.3 Thus the explanatory terminology I propose distinguishes between a theorised agent-image, the psychological and cognitive mechanisms that may affect agents' actions as well as the social mechanisms that contribute both to actors' actions and their joint outcomes. What this means, in terms of elaboration, is that not only is an account of the mechanism(s) operating in a situation needed, but also an account of the agent-image employed is required. Furthermore, an account of the relevant contexts of action, something that is inherent in the actionalism thesis' insistence on describing the actors' situation, is also required. In an illuminating paper on the use of EbM in political analysis, Falleti and Lynch argue that 'credible causal social scientific explanation can occur if and only if researchers are attentive to the interaction between causal mechanisms and the context in which they operate' (Falleti and Lynch 2009, 2). The contextual element is particularly important in comparative historical sociology (see also Mjøsset 2011).

6.4 Walter Brustein's book *The Logic of Evil* (1996) is a fairly well-known example of rational action sociology applied to a historical topic. It is also an example of the explanatory importance of contexts. Brustein's topic is the support of the Nazi party in Germany before Hitler took power. In the beginning of the book Brustein reviews and shows as questionable a number of interpretations of why the Nazis were, over the years, supported by Germans from various social categories, ranging from farmers to businessmen. Brustein's chosen framework for studying the problem is that of a rational voter: people supported the Nazi party because they saw the party as defending their interests. However, the agent-image chosen plays a quite small role in the book. Most of it is an empirically backed investigation into the economic, social and political situation of various social groups and their propensity to join the Nazi party and remain as its members. Brustein found the Nazi party's populist programme attracted people from various socioeconomic environments. But, if their situation changed they lost interest and did not renew their membership. Thus the agent-image was the 'constant' that made the contextual 'variables' explanatory.

A Summary and some Conclusions

7.1 The whole of the discussion presented here has been conducted in terms of social scientific (sociological) explanation. The idea of EbM is, of course, not restricted to sociology as other human sciences as well as many 'hard' sciences employ the mechanism metaphor, especially biology. Moreover, a great part of what is written about EbM by *empirically oriented* sociologists is research using large-N data sets that are analysed with advanced statistical methods. However, the idea of EbM has also been hotly debated by comparative-historical sociologists whose research often combines qualitative and simple statistical analysis. I do not see EbM only as a corrective for statistical 'variables sociology'-- subject to a lot of critique nowadays -- focusing on correlations and disregarding their elaboration. Instead, I also see it as a more general strategy of explanation that is connected with all kinds of sociology, whether quantitative, qualitative, or both.

7.2 *When deconstructed, EbM can be seen as entailing two or even three distinct theses.* The first one, the elaboration thesis, does not require the idea of mechanism as one of its elements. It simply says that observed or inferred interdependencies between factors (events, variables) do not qualify as a full explanation of those interdependencies, unless a plausible and verifiable account of how those interdependencies come about can be provided. This I consider a very important point for anyone trying to explain something or trying to assess an explanation provided by someone else. However, there is no compelling reason why that kind of elaboration should be constructed in terms of the notion of mechanism. In fact, it is possible that the use of the metaphor of mechanism might even seduce us into the wrong kind of *Gedankenbilder* about society (see Gorski 2009, 151), such as social systems as well as actors function machine-like, mechanically. Should we then abandon the term 'mechanism' and invent a new terminology? Due to the well-established nature of the terminology in the social sciences, I am not ready to do that. Instead, I have sketched a terminology where social and *other* mechanisms have a more specific, albeit limited, meaning than they have previously had.

7.3 I argued above that accounting for the mechanisms operative in the situation is not sufficient for explanation. A basic point made by both critical realists and analytical sociologists is that whether a mechanism is switched on or not, depends on the situation. Accounting for the context where the supposed mechanism should operate becomes important for the explanation. Moreover, the proposed terminology limits the concept of mechanism to what has an effect on actors' actions and behaviours and to how that affects the unintended consequences of their outcomes. On one hand, there are psychological/cognitive mechanisms, on the other hand, there are social mechanisms that are constituted by actors' actions and interactions, and I consider it important to distinguish between these two. Apart from hypotheses about mechanisms and propositions about relevant contexts one needs an agent-image:

a theoretical conception about how human action typically comes about, regarding, for example, whether actors are moved by the rational pursuit of self-interest or conform to habitual behaviour patterns learned in their socialisation process. Different agent-images produce different kinds of elaborations, as well as different kinds of ideas about possible mechanisms.

7.4 I am not the first to point that EbM does not in any way imply the third thesis, that of actionalism. It is in principle possible to provide elaborations that are not constructed by taking actors and their actions as the starting point – elaborations that operate, for example, on the emergent properties of social structures. However, *it is rather hard to see that non-actionalist elaborations could really provide a causal understanding of a situation*. The problem is that society happens through people's actions and interactions, and to make sense of them one cannot avoid accounting for how and why people act or behave as they do. A purely structuralist explanation is defective unless it can provide an account of how structures do indeed determine the actors' action to the extent that actors' motivations and choices need not be accounted for at all. In fact, a clever structuralist does not deny that society happens through people's choices but he or she does not grant actors' choices any explanatory importance. Actions can sufficiently be explained by examining the actors' structural locations, but how structure produces action still remains to be adequately demonstrated. So, my conclusion is that though not implied by EbM, the actionalism thesis must indeed be accepted if we really want to elaborate our explanations.

7.5 There is also a point I want to make that goes beyond the discussion of explanation in this paper. Long ago, Jon Elster indicated that thinking in terms of social and other kinds of mechanisms just might be a more fruitful way to advance social science than trying to develop universal theories of social behaviour (see Elster 1993, 2). Elster's argument is that social researchers' quest for universal social laws is a vain one, and they would do better to try to further develop and refine the repertoire of social mechanisms that the social sciences have discovered. This point applies not only to social research having as its explicit goal the construction of causal explanations. Many want to restrict causal explanation to quantitative social science only (though Elster does not) and assign to qualitative social science tasks other than explanation, for example, interpretation or a theoretically-laden description of the research object. In fact, also practitioners of qualitative or historical-comparative sociology sometimes do that, often thinking of causal explanation as belonging to the natural sciences rather than the social sciences. The main question of qualitative research (from the sociological perspective) is often: 'What is going on here?' The answer to this question might very well involve an account of not only the actions and interactions of actors but also the social and other mechanisms that affect the actors' actions – and the latter description just might be the real social scientific contribution of the research. Whatever the standpoint towards the role of causal explanation is in sociology, the strategy of seeking mechanisms remains a fruitful option for which a particular programme of sociology can be constructed.

Notes

¹The critical realists' view of explanation was originally inspired by Rom Harré's realist philosophy of science (see Harré 1970). Harré is not a critical realist social scientist as he is very suspicious of anything more grand than the micro-sociology of action whereas structural sociology is the bread and butter of critical realist sociology.

²Hence, it is astonishing that the critical realist Peter Manicas (2006, 81) has praised Gudmund Hernes' analysis of social mechanisms as an exemplary one.

³Philip Gorski has presented similar kinds of reservations about the metaphor of mechanism (see Gorski 2009, 154-156).

Acknowledgements

The paper was drafted while working as a senior researcher of Academy of Finland (Grant 111731) and it was finished while directing the research project Neighbouring and neighbor disputes in contemporary society (Grant 127943). I thank Academy of Finland for economic support for my research. I also thank three anonymous referees as well as many colleagues, e.g. in the 24th Conference of the Nordic Sociological Association (Århus 2008) and Conference on Analytical Sociology and Social Mechanisms (Barcelona 2010), for useful comments on this paper.

References

- ABBOTT, A (2007) 'Mechanisms and Relations', *Sociologica*, 2/2007:
<<http://www.sociologica.mulino.it/journal/article/index/Article/Journal:ARTICLE:105>>.
- BENGTTSSON, B (1998) 'Tenants' dilemma – on collective action in housing', *Housing Studies* Vol. 13, Issue 1 pp. 99–120. [doi:[dx.doi.org/10.1080/02673039883515](https://doi.org/10.1080/02673039883515)]
- BENGTTSSON, B (2001) 'Solving the Tenants' Dilemma: Collective Action and Norms of Co-operation in Housing', *Housing, Theory and Society* Vol. 17, Issue 4 pp. 175–187.
- BHASKAR, R (1979) *The Possibility of Naturalism: A philosophical critique of the contemporary human sciences*. Brighton: The Harvester Press.

- BHASKAR, R (1986) *Scientific Realism and Human Emancipation*. London: Verso.
- BROWN, R H (1977) *A Poetic for Sociology. Toward a Logic of Discovery for the Human Sciences*. Chicago and London: The University of Chicago Press.
- BRUSTEIN, W (1996) *The Logic of Evil. The Social Origins of the Nazi Party 1925-1933*. New Haven and London: Yale University Press.
- COHEN, G A (1980) Functional Explanation: Reply to Elster. *Political Studies* Vol. XXVIII, Issue 1, pp. 129-135. [doi://dx.doi.org/10.1111/j.1467-9248.1980.tb01240.x]
- ELSTER, J (1989a) *The Cement of Society. A Study of Social Order*. Cambridge, New York, Port Chester, Melbourne, Sydney: Cambridge University Press.
- ELSTER, J (1989b) *Nuts and Bolts for the Social Sciences*. Cambridge, New York, Port Chester, Melbourne, Sydney: Cambridge University Press.
- ELSTER, J (1993) *Political Psychology*. Cambridge, New York, Oakleigh: Cambridge University Press.
- ELSTER, J (1998) 'A plea for mechanisms' in Hedström P and Swedberg R (eds.) *Social Mechanisms. An analytical approach to social theory*. Cambridge: Cambridge University Press. [doi://dx.doi.org/10.1017/CBO9780511663901.003]
- FALLETI, T G and Lynch J F (2009) 'Context and causal mechanisms in political analysis', *Comparative Political Studies* Vol. 42, Issue 9 pp. 1143–1166. [doi://dx.doi.org/10.1177/0010414009331724]
- GAMBETTA, D (1998) 'Concatenations of mechanisms' in Hedström, P and Swedberg, R (eds.) *Social Mechanisms. An analytical approach to social theory*. Cambridge: Cambridge University Press. [doi://dx.doi.org/10.1017/CBO9780511663901.005]
- GORSKI, P (2009) 'Social "Mechanisms" and Comparative-Historical Sociology' in Hedström P and Wittrock B (eds.) *Frontiers of Sociology*. Leiden and Boston: Brill. [doi://dx.doi.org/10.1163/ej.9789004165694.i-450.48]
- GROSS, N (2009) 'A Pragmatist Theory of Social Mechanisms', *American Sociological Review* Vol. 74, Issue 3 pp. 358-379. [doi://dx.doi.org/10.1177/000312240907400302]
- GUSFIELD, J R (1986) *Symbolic Crusade. Status Politics and the American Temperance Movement. Second Edition*. Urbana and Chicago: University of Illinois Press.
- HARRÉ, R (1970) *The Principles of Scientific Thinking*. Chicago: University of Chicago Press.
- HARRÉ, R (1988) 'Modes of Explanation', in Hilton, D J (ed.) *Contemporary Science and Natural Explanation*. Brighton: The Harvester Press.
- HARRÉ, R, Clarke, D and De Carlo, N (1985) *Motives and Mechanisms. An introduction to the psychology of action*. London and New York: Methuen.
- HEDSTRÖM, P (2005) *Dissecting the Social. On the Principles of Analytic Sociology*. Cambridge: Cambridge University Press.
- HEDSTRÖM, P and Bearman, P (2009) 'What is Analytical Sociology All About? An Introductory Essay' in Hedström P and Bearman P (eds.) *The Oxford Handbook of Analytical Sociology*. Oxford: Oxford University Press.
- HEDSTRÖM, P and Swedberg, R (eds.) (1998) *Social Mechanisms. An analytical approach to social theory*. Cambridge: Cambridge University Press.
- HEDSTRÖM, P and Swedberg, R (1998) 'Social Mechanisms: An introductory essay' in Hedström, P and Swedberg, R (eds.) (1998).
- HEDSTRÖM, P and Ylikoski, P (2010) 'Causal Mechanisms in the Social Sciences', *Annual Review of Sociology* Vol. 36 pp. 49-67.
- HERNES, G (1998) 'Real virtuality', in Hedström, P. & Swedberg, R. (eds.) *Social Mechanisms. An analytical approach to social theory*. Cambridge: Cambridge University Press. [doi://dx.doi.org/10.1017/CBO9780511663901.004]
- HUNT, A (1999) 'Anxiety and social explanation: some anxieties about anxiety', *Journal of Social History* Vol. 32, Issue 3 pp. 509-528. [doi://dx.doi.org/10.1353/jsh/32.3.509]
- KEAT, R and Urry, J (1982) *Social theory as science. 2nd edition*. London: Routledge and Kegan Paul.
- KISER, E and Hechter, M (1991) 'The Role of General Theory in Comparative-Historical Sociology', *American Journal of Sociology* Vol. 97, Issue 1 pp. 1-30. [doi://dx.doi.org/10.1086/229738]
- LLOYD, C (1986) *Explanation in Social History*. Oxford: Blackwell.

- MAHONEY, J (2001) 'Beyond Correlational Analysis: Recent Innovations in Theory and Method', *Sociological Forum* Vol. 16, Issue 3 pp. 575-593. [doi://dx.doi.org/10.1023/A:1011912816997]
- MANICAS, P (2006) *A Realist Philosophy of Social Science. Explanation and Understanding*. Cambridge: Cambridge University Press. [doi://dx.doi.org/10.1017/CBO9780511607035]
- MANZO, G (2010) 'Analytical Sociology and Its Critics', *Archives Européennes de Sociologie* Vol. LI, Issue 1 pp. 129-170.
- MERTON, R K (1968) *Social Theory and Social Structure. 1968 Enlarged Edition*. New York: The Free Press.
- MJÖSET, L (2011) 'Review Essay', *Acta Sociologica* Vol. 54, Issue 3 pp. 301-305.
- OUTHWAITE, W (1987) *New Philosophies of Social Science: Realism, hermeneutics and critical theory*. Basingstoke: Macmillan.
- PAWSON, R (1989) *A Measure for Measures. A manifesto for empirical sociology*. London and New York: Routledge.
- POPPER, K-R (1979) *Objective Knowledge. An Evolutionary Approach. Revised edition*. London: Clarendon Press.
- QUADAGNO, J and Knapp, S (1992) 'Have Historical Sociologists Forsaken Theory? Thoughts on History/Theory Relationship', *Sociological Methods and Research* Vol. 20, Issue 4 pp. 481-507. [doi://dx.doi.org/10.1177/0049124192020004004]
- RUNCIMAN, W G (2000) *The Social Animal*. Ann Arbor: The University of Michigan Press.
- RÄSÄNEN, P (2003) *In the Twilight of Social Structures. A mechanism-based study of contemporary consumer behaviour*. Annales Universitatis Turkuensis B 263. Turku: University of Turku.
- SAYER, A (1984) *Method in Social Science: A realist approach. 2nd edition*. London: Hutchinson. [doi://dx.doi.org/10.4324/9780203310762]
- SHELLING, T (1971) 'Dynamic Models of Segregation', *Journal of Mathematical Sociology* Vol. 1, Issue 2 pp. 143-186. [doi://dx.doi.org/10.1080/0022250X.1971.9989794]
- SEN, A (2006) *Identity and Violence. The illusion of destiny*. New York: W. W. Norton and Co.
- SOMERVILLE, P and Bengtsson, B (2002) 'Constructionism, Realism and Housing Theory', *Housing, Theory and Society* Vol. 19, Issues 3-4 pp. 121-136.
- VAN PARIJS, P (1981) *Evolutionary Explanation in the Social Sciences. An emerging paradigm*. Totowa: Rowman and Littlefield.
- WALLIS, R and Bruce, S (1986) *Sociological Theory, Religion and Collective Action*. Belfast: The Queen's University.
- WALLIS, R (1979) *Salvation and Protest. Studies of Social and Religious Movements*. New York: St. Martin's Press.